A DIRECT DETERMINATION OF THE MINIMUM AREA BETWEEN A CURVE AND ITS CAUSTIC.

BY OTTO DUNKEL.

If descending parallel rays of light lying in a plane fall upon a curve in that plane, a caustic* will be produced by the portions of the curve whose concavity is toward the light. The remaining portions produce no actual caustic but a virtual caustic, which becomes an actual caustic by reversing the direction of the rays. Both the actual and the virtual caustic, if any, appear in the analytical treatment. Given two points, the origin and the point $P_2(x_2, y_2)$ in the first quadrant, and a curve joining these two points with given inclinations at the points τ_1, τ_2 such that $-\pi/2 < \tau_1 < \tau_2 < \pi/2$, the area S enclosed by the curve, its caustic and the reflected rays at the two points will be considered, and the form of the curve will be determined which makes this area a minimum. Several cases of end conditions will be examined. This problem is easily treated by the methods of the Calculus of Variations,† but a more elementary method will be used here which seems to be better adapted to this special form of minimum problem, as it yields the results more directly and rapidly. The method applies in precisely the same manner to the problem of the minimum area between a curve and its evolute. It will be seen that this method is quite analogous to the high-school algebra method of solving problems of maxima and minima of quadratic functions by the device of completing the square of the quadratic function.

The Determination of the Minimizing Curve and the Area. If R is the radius of curvature of the curve at a point P at which the inclination is τ and K is the point of contact of the reflected ray with the caustic, then

$$(1) \qquad \delta = \frac{R\cos\tau}{2} = \frac{1}{2}\frac{ds}{d\tau}\cos\tau = \frac{1}{2}\frac{dx}{d\tau},$$

where $\delta = PK$ is positive when the concavity of the curve is upward and negative when downward.‡ The element of area between two reflected

* Dunkel, "Note on caustics," The American Mathematical Monthly, vol. XXVII, 1920, no. 5, page 225.

† Dunkel, "The curve which with its caustic encloses the minimum area," Washington University Studies, Scientific Series, vol. VIII, No. 2, pp. 183–194, Jan. 1921.

‡ If P and P' are neighboring points on the curve, let the normals at these two points meet in C and the reflected rays in Q. Let the perpendicular to $P'C$ at its middle point meet PC in M'; then P, P', Q and M' lie upon a circle, since $\angle PQP' = 2\angle PCP' = \angle PM'P'$. When P' approaches P, the limiting position of M' is M, the middle point of the radius of curvature R at

rays is $\delta \cos \tau \, ds/2$, which becomes $\delta^2 d\tau$ by use of (1). The integral to be made a minimum is then

$$(2) \qquad S = \int_{\tau_1}^{\tau_2} \delta^2 d\tau,$$

with the two auxiliary conditions obtained from (1),

$$(3) \qquad x_2 = 2\int_{\tau_1}^{\tau_2} \delta d\tau, \qquad y_2 = 2\int_{\tau_1}^{\tau_2} \delta \tan \tau \, d\tau.$$

It will be assumed that the curves considered are such that δ is a continuous function of τ. Let A and B denote two constants which will be determined later, then, after multiplying the first equation in (3) by $B/2$ and the second by $A/2$, the integral (2) may be written

$$S = \int_{\tau_1}^{\tau_2} [\delta^2 - (A \tan \tau + B)\delta] d\tau + \frac{1}{2}(Ay_2 + Bx_2).$$

This suggests the transformation to the form

$$(4) \quad S = \int_{\tau_1}^{\tau_2} \left[\delta - \left(\frac{A \tan \tau + B}{2}\right)\right]^2 d\tau - \int_{\tau_1}^{\tau_2} \left(\frac{A \tan \tau + B}{2}\right)^2 d\tau + \frac{1}{2}(Ay_2 + Bx_2).$$

Hence the minimum value of S will be given by

$$(5) \qquad \delta = \frac{1}{2}(A \tan \tau + B),$$

provided that the constants A and B can be chosen uniquely so that δ satisfies the conditions (3). If this can be done, the equality (4) gives the expression for the minimum area

$$(6) \qquad S = \frac{1}{4}(Ay_2 + Bx_2).$$

The constants A and B are to be determined from the equations resulting from (3),

$$(7) \qquad \begin{aligned} x_2 &= A\int_{\tau_1}^{\tau_2} \tan \tau \, d\tau + B\int_{\tau_1}^{\tau_2} d\tau, \\ y_2 &= A\int_{\tau_1}^{\tau_2} \tan^2 \tau \, d\tau + B\int_{\tau_1}^{\tau_2} \tan \tau \, d\tau; \end{aligned}$$

and there exists a unique solution of these equations if their determinant

P, the limit circle has PM as a diameter and it cuts the reflected ray PQ in K, a point on the caustic. Thus $PK = \delta = R \cos \tau/2$, where $\tau = \angle MPK$. This result may also be obtained from the general formula given in The American Mathematical Monthly, l. c. Another derivation is given in the Washington University Studies, l. c.

D_2 is not zero. But this determinant is the negative of the discriminant of the quadratic form in A and B, regarded as variables,

$$A^2 \int_{\tau_1}^{\tau_2} \tan^2 \tau \, d\tau + 2AB \int_{\tau_1}^{\tau_2} \tan \tau \, d\tau + B^2 \int_{\tau_1}^{\tau_2} d\tau = \int_{\tau_1}^{\tau_2} (A \tan \tau + B)^2 d\tau.$$

Since this form is never negative and vanishes only when *both* A and B are zero, and the coefficients of the squared terms are positive, its discriminant must be greater than zero, and hence D_2, the determinant of the equations (7), must be less than zero for $\tau_2 \neq \tau_1$. For $\tau_2 = \tau_1$ it is clear that $D_2 = 0$. The constants A and B can, therefore, be determined uniquely, and hence the value of δ in (5) gives the minimum area. It may be observed that if η indicates the variation from δ as given in (5), then equation (4) may be written

$$\Delta S = \int_{\tau_1}^{\tau_2} \eta^2 \, d\tau, \tag{4'}$$

where ΔS denotes the increment of the area due to the variation η. The parametric equations of the curve are obtained by integration from equations similar to (7) in which x_2, y_2, τ_2 are replaced by x, y, τ, respectively, and it will be found that

$$\begin{aligned} x &= A \log\left(\frac{\sec \tau}{\sec \tau_1}\right) + B(\tau - \tau_1), \\ y &= A\left[\tan \tau - \tan \tau_1 - (\tau - \tau_1)\right] + B \log\left(\frac{\sec \tau}{\sec \tau_1}\right). \end{aligned} \tag{8}$$

From (1) and (5) follow the equations

$$\frac{dx}{d\tau} = A \tan \tau + B, \qquad \frac{dy}{d\tau} = (A \tan \tau + B) \tan \tau, \tag{8'}$$
$$R = (A \tan \tau + B) \sec \tau,$$

which are useful in the study of the appearance of the curve. If $A \neq 0$, they show that the curve has a cusp at the point for which $\tan \tau_0 = -B/A$. If τ_0 lies between τ_1 and τ_2, then, since δ changes sign, there is a virtual caustic given by the part of the curve for which the inclination is greater than τ_0. A discussion of the properties of the curve and of its caustic has been given in another paper,* and it is there shown that the caustic has, in general, two cusps with tangents parallel to that of the cusp of the original curve.

A Cusp at One End Point. If the minimizing curve is such that $\delta = 0$ at an end point, there is a cusp at that end and the curve is concave up from the initial point to the other end. For convenience it will be assumed

* Washington University Studies, l.c.

that the cusp is at P_2. Let τ_0 denote the inclination at this point, A_0, B_0, and δ_0 indicate the determinations of A, B, and δ for this case, so that

$$\delta_0 = \frac{1}{2}(A_0 \tan \tau + B_0).$$

If we consider any other curve passing through the origin with the inclination τ_1, and through P_2 with the inclination $\tau_2 > \tau_1$ and such that its δ does not change sign from τ_1 to τ_2, then the area given by this curve is greater than the area given by the curve δ_0. If $\tau_2 = \tau_0$, the truth of the statement follows from the previous work, so we may assume now that $\tau_2 \neq \tau_0$. For the curve δ the two equations (3) must be satisfied, while for the minimizing curve δ_0 similar equations must be written in which τ_2, δ are replaced by τ_0, δ_0. From these four equations follow the pair of equations

$$(9') \qquad \int_{\tau_1}^{\tau_2} \delta d\tau = \int_{\tau_1}^{\tau_0} \delta_0 d\tau, \qquad \int_{\tau_1}^{\tau_2} \delta \tan \tau \, d\tau = \int_{\tau_1}^{\tau_0} \delta_0 \tan \tau \, d\tau.$$

Multiplying the first equation by $B_0/2$ and the second by $A_0/2$ and adding the corresponding sides, it will be found that

$$(9) \qquad \int_{\tau_1}^{\tau_2} \delta\delta_0 \, d\tau = \int_{\tau_1}^{\tau_0} \delta_0^2 \, d\tau.$$

Comparing the two areas, we have

$$(10) \qquad \begin{aligned} \Delta S = \int_{\tau_1}^{\tau_2} \delta^2 \, d\tau - \int_{\tau_1}^{\tau_0} \delta_0^2 \, d\tau &= \int_{\tau_1}^{\tau_2} (\delta - \delta_0)^2 d\tau + 2 \int_{\tau_1}^{\tau_2} \delta\delta_0 \, d\tau \\ &\qquad - 2\int_{\tau_1}^{\tau_0} \delta_0^2 \, d\tau - \int_{\tau_0}^{\tau_2} \delta_0^2 \, d\tau \\ &= \int_{\tau_1}^{\tau_2} (\delta - \delta_0)^2 d\tau + \int_{\tau_2}^{\tau_0} \delta_0^2 \, d\tau, \end{aligned}$$

in which the last line follows by use of (9). Hence if $\tau_2 \leqq \tau_0$, $\Delta S > 0$ and the theorem is true in this case. If $\tau_2 > \tau_0$, it will be of aid to write (10) in the following form:

$$(10') \qquad \Delta S = \int_{\tau_1}^{\tau_0} (\delta - \delta_0)^2 d\tau + \int_{\tau_0}^{\tau_2} \delta^2 \, d\tau - 2\int_{\tau_0}^{\tau_2} \delta\delta_0 \, d\tau,$$

which shows that ΔS is again greater than zero, for the last integral on the right is negative since $\delta > 0$ and $\delta_0 \leqq 0$ from τ_0 to τ_2. The reasoning fails in this second part if the comparison curve is allowed to have a cusp through the change of sign of δ, and in what follows it will be shown how to find curves of this kind giving a smaller area than δ_0.

The Minimum Area as a Function of the End Inclinations. Suppose now that two minimizing curves, δ_1, δ_2, passing through the given end points have the same inclination τ_1 at the origin and the inclinations τ_2' and τ_2'', respectively, at P_2, and let the respective minimum areas be S_1 and S_2. Then $S_1 > S_2$ if $\tau_2' < \tau_2''$. This follows at once from the formula

$$\Delta S = \int_{\tau_1}^{\tau_2''} \delta_2^2 \, d\tau - \int_{\tau_1}^{\tau_2'} \delta_1^2 \, d\tau = - \int_{\tau_1}^{\tau_2'} (\delta_2 - \delta_1)^2 d\tau - \int_{\tau_2'}^{\tau_2''} \delta_2^2 \, d\tau$$

which is obtained in identically the same way as (10), and, since no use is made of the fact that δ_1 is a minimizing curve, i.e., that

$$\delta_1 = (A_1 \tan \tau + B_1)/2,$$

it shows that the minimizing curve δ_2 gives a smaller area than *any other* curve through the same end points and having the same initial inclination but a final inclination less than or equal to that of the minimizing curve. A slightly different proof will also be given as it leads to an additional interesting result. Making use of the facts that $\delta_1 = (A_1 \tan \tau + B_1)/2$ and $\delta_2 = (A_2 \tan \tau + B_2)/2$, two equations similar to (9) may be written

$$\begin{aligned} S_1 &= \int_{\tau_1}^{\tau_2'} \delta_1^2 \, d\tau = \int_{\tau_1}^{\tau_2''} \delta_1 \delta_2 \, d\tau, \\ S_2 &= \int_{\tau_1}^{\tau_2''} \delta_2^2 \, d\tau = \int_{\tau_1}^{\tau_2'} \delta_1 \delta_2 \, d\tau. \end{aligned} \tag{11}$$

Hence

$$\Delta S = S_2 - S_1 = - \int_{\tau_2'}^{\tau_2''} \delta_1 \delta_2 \, d\tau. \tag{12}$$

Referring to the determination of A and B in (7) it will be seen that these two functions of τ_2 are continuous in τ_2 as long as $\tau_2 \neq \tau_1$. Hence δ is a continuous function of τ and τ_2, and it follows that, by taking $\tau_2'' - \tau_2'$ small enough, the sign of δ_2 can be made the same as that of δ_1 for all values of τ in the interval of integration, if $\delta_1 \neq 0$ for $\tau = \tau_2'$. It follows then that ΔS is negative. Since S is a continuous function of τ_2, which appears in the integrand as well as in the upper limit, it follows that S decreases even at points for which $\delta = 0$. The equation (12) leads by a simple reasoning to the result

$$\frac{dS}{d\tau_2} = - \delta^2 \tag{12'}$$

which also shows that S decreases as τ_2 increases. A similar analysis may be applied to the other extremity with the result that the area S decreases as τ_1 decreases.

The Symmetric Solution. If the two end points are taken on the same level, say $y_2 = 0$, and the inclinations at the ends are taken as the negatives of each other, $\tau_2 = -\tau_1 > 0$, then (7) shows that $A = 0$ and the equation (8) reduces to

$$(13) \qquad y = B \log \frac{\sec\left(\dfrac{x}{B} + \tau_1\right)}{\sec \tau_1} \qquad \begin{aligned} x_2 &= B(\tau_2 - \tau_1), \\ \delta &= B/2. \end{aligned}$$

In this case the curve is without a cusp and δ is a constant; and thus its caustic has a property somewhat similar to the tractrix. This curve is called the catenary of uniform strength. This is a solution of the problem in which the end conditions may be stated as follows. Given two vertical straight lines at $x = 0$ and $x = x_2$ and curves crossing these lines with the inclinations τ_1, τ_2, respectively, then (13) is the curve which gives the minimum area enclosed between it, its caustic, and the reflected rays at the crossing points. From the nature of the problem it will be seen that it is no real restriction to assume that all the curves pass through the origin. In this case the equation for y_2 in (3) drops out and the equation (4) becomes

$$S = \int_{\tau_1}^{\tau_2} \left(\delta - \frac{B}{2}\right)^2 d\tau - \int_{\tau_1}^{\tau_2} \left(\frac{B}{2}\right)^2 d\tau + \frac{B}{2} x_2,$$

and the minimum area is given by $\delta = B/2$ and has the value $Bx_2/4$, where B is to be determined from the equation for x_2 in (3).

This result may also be obtained by determining the value of y_2 which makes the minimum area in (6) attain its least value. Solving the equations (7) or (8) for A and B, we have

$$(14) \quad 4S = Ay_2 + Bx_2 = \frac{\left[y_2(\tau_2 - \tau_1) - x_2 \log \dfrac{\sec \tau_2}{\sec \tau_1}\right]^2}{-(\tau_2 - \tau_1)D_2} + \frac{x_2^2}{(\tau_2 - \tau_1)}.$$

Remembering that D_2 is negative it is clear that if

$$(15) \qquad y_2(\tau_2 - \tau_1) - x_2 \log \frac{\sec \tau_2}{\sec \tau_1} = 0,$$

S reaches its minimum value $x_2^2/4(\tau_2 - \tau_1)$. But the expression to the left in (15) is the value of $-D_2A$ and hence $A = 0$ gives the minimum area S.